

INSIDE THIS WEEK: TECHNOLOGY QUARTERLY

The Economist

MARCH 7TH-13TH 2009

Economist.com

Obama's flawed tax plan
Pakistan, cricket and terrorism
Bringing justice to Sudan
How low can dividends go?
The bees are back in town

How to stop the drug wars



US\$6.99 - C\$7.99



Argentina.....\$7.00
Bahamas.....\$9.95
Barbados.....Bds\$16.50
Bermuda.....Bds\$7.00
Brazil.....R\$29.90

Canada.....C\$7.99
Chile.....Ch\$5,000
Colombia.....Cof\$22,000
Costa Rica.....C\$300
Guyana.....GYD 1,450

Jamaica.....J\$510
Mexico.....Mex\$70
Peru.....S/ 36.00
Spain.....€3.50
St Maarten.....\$9.25

Trinidad & Tobago.....TDS 43
Turks & Caicos.....\$9.50
UK.....£4.00
USA.....US\$6.99
Venezuela.....Bs27

The internet's librarian

Avec plaisir!
Jeanele Borg

Brewster Kahle wants to create a free, online collection of human knowledge. It sounds impossibly idealistic—but he is making progress

FOR a man who has set himself a seemingly impossible mission, Brewster Kahle seems remarkably laid back. Relaxing in the black leather recliner that serves as his office chair, his stockinged feet wriggling with evident enthusiasm, the founder of the Internet Archive explains what has driven him for more than a decade. "We are trying to build Alexandria 2.0," says Mr Kahle with a wide-eyed, boyish grin. Sure, and plenty of people are trying to abolish hunger, too.

It would be easy to dismiss Mr Kahle as an idealistic fruitcake, but for one thing: he has an impressive record when it comes to setting lofty goals and then lining up the people and technology needed to get the job done. "Brewster is a visionary who looks at things differently," says Carole Moore, chief librarian at the University of Toronto. "He is able to imagine doing things that everyone else thinks are impossible. But then he does them."

Mr Kahle is an unostentatious millionaire who does not "wear his money on clothes", as one acquaintance graciously puts it. But behind his dishevelled demeanour is a skilled technologist, an ardent activist and a successful serial entrepreneur. Having founded and sold technology companies to AOL and Amazon, he has now devoted himself to building a non-profit digital archive of free materials—books, films, concerts and so on—to rival the legendary Alexandrian library of antiquity. This has brought him into conflict with Google, the giant internet company which is pursuing a similar goal, but in a rather different (and more commercially oriented) way.

Biblio-tech

After graduating in 1982 from the Massachusetts Institute of Technology (MIT), where he had studied with Marvin Minsky, an artificial-intelligence guru, Mr Kahle joined a group of MIT alumni who were founding a company, Thinking Machines, that made parallel supercomputers. There Mr Kahle worked alongside such luminaries as Richard Feynman (a Nobel prize-winning American physicist), Dr Minsky and Daniel Hillis, a maverick computer scientist best known as the inventor of the 10,000-year clock.

Building on the search technology developed at Thinking Machines, Mr Kahle left to found his own company, WAIS Inc, in 1989. It took its name from the Wide Area Information Server protocol, an early form of internet search engine which had been developed by Thinking Machines with Apple, Dow Jones and KPMG, and made software for online publishing. Its customers included the *Wall Street Journal*, which was setting up the first subscription-based online news site, and CMP, a magazine company that pioneered internet advertising. Mr Kahle was a decade ahead of his competitors in grasping the importance of payment systems, online privacy and user ratings. AOL bought the firm in 1995 for an undisclosed sum, thought to be around \$15m.

Mr Kahle—who by 1996 had almost a dozen patents to his name—quickly turned to his next project. He founded the non-profit Internet Archive and, with a former colleague, co-founded a firm called Alexa that tracks and analyses the paths people follow as they move around the web, in order to direct people with similar interests to relevant information. Amazon bought Alexa for an estimated \$250m in 1999. Mr Kahle continued to work on Alexa until 2002, but then dedicated himself fully to the Internet Archive.

The most famous part of the archive is the Wayback Machine (its name inspired by the WABAC machine in the 50-year-old television cartoon featuring Rocky and Bullwinkle). This online attic of digital memorabilia stores copies of internet sites so that people can see, for example, what economist.com looked like in January 1997. Paul Courant, the dean of libraries at the University of Michigan, equates what the archive does for the internet with what the British Museum did for the British empire. "The internet has become the medium of choice for a great deal of cultural production," he says. The Wayback Machine "gives us access to what people were producing at different points in time," he says. Evidently this is of more than just academic interest: the site gets 500 page requests per second.

In addition to this archive of web pages there is also an audio library with more than 300,000 MP3 files, a moving-images archive with more than 150,000 films and videos, and a live-music archive with recordings of more than 60,000 concerts. All the collections are available free to anyone with internet access, each gathering its own set of fans. A remarkably popular archive within the audio library is ▶



It is easy to dismiss Mr Kahle as an idealist, but he has an impressive record of getting things done.



► devoted to the Grateful Dead.

But all these things are steps towards Mr Kahle's wider goal: to build the world's largest digital library. He has recruited 135 libraries worldwide to openlibrary.org, the aim of which is to create a catalogue of every book ever published, with links to its full text where available. To that end, the Internet Archive is also digitising books on a large scale on behalf of its library partners. It scans more than 1,000 books every day, for which the libraries pay about \$30 each. (The digital copy can then be made available by both parties.)

Some 200 people work for the Internet Archive, which has an annual budget of \$10m-14m. Initially funded by Mr Kahle, the archive now gets much of its income from grants made by foundations and from libraries that pay it to digitise their books. It also runs a variety of one-off projects, such as a collaboration with America's space agency, NASA, to make available photos and films relating to the history of the space programme, and a "print on demand" system to turn digital files into physical books in minutes.

With his happy-go-lucky management style, Mr Kahle comes across as easy-going. But the 48-year-old has been known to stand his ground—even against the tough guys. "Come back when you have a warrant," reads the floor mat underneath his office recliner. It was a gift from the Electronic Frontier Foundation (an activist group on whose board Mr Kahle sits) after Mr Kahle refused to hand over information about one of the Internet Archive's users to the Federal Bureau of Investigation in 2007.

This activist for online privacy is also a staunch supporter of openness. He insisted that the Internet Archive's specially developed scanning machine, called Scribe, should be an open-source device, meaning that details of how it works are made available to anyone who wants them. The same is true of the "PetaBox", another archive-developed machine that holds up gigabytes of data. "Everything Brewster does is open. He personifies openness," says John Seely Brown, who sits on Amazon's board of directors and was previously the chief scientist at Xerox, and the director of its Palo Alto Research Centre. Being open "is the right way to have a thriving industry," says Mr Kahle. "I have been much more successful when letting people know what I want to do. I get much more help that way."

Underlying Mr Kahle's enthusiasm for openness is an implicit criticism of the

much larger book-scanning project being undertaken by Google. Like Mr Kahle, Google's founders have a lofty goal: "to organise the world's information and make it universally accessible and useful." Since much of the world's information is in books, this means large-scale scanning. But whereas Mr Kahle has focused on old books that are no longer protected by copyright, and making the full text available, Google's Book Search project has scanned some 7m more recent works, most of them still covered by copyright, and allows access only to small chunks.

Google argued that since it was not making entire works available, it was not infringing copyright and did not need permission from publishers to display these small chunks (with advertising alongside them). The publishing industry disagreed and sued Google, and a settlement was reached in October 2008. It is still subject to a judge's approval, but could be finalised by June. Under the terms of the settlement, Google will put copyrighted works online only with the permission of publishers, who can also decide whether to make a preview available or not. Google will also be allowed to sell access to entire books online, sharing the proceeds with publishers. It has, in other words, struck a deal that will allow it to go on scanning books and make money providing access to them online.

Mr Kahle's approach to broadening the number of books available for his archive was rather different. He unsuccessfully sued the American government, in a case known as *Kahle v Gonzales*, in an effort to roll back what he regards as excessive copyright terms. Reducing the period of copyright protection would have dramatically expanded the universe of copyright-free works, and hence the number that could be scanned and made available online. This would have benefited everyone—not just Mr Kahle and his project.

Literary criticism

Google's legal settlement has caused controversy because it means that Google is now the only big company to be building a significant digital collection of copyrighted books. Some librarians worry that this gives the internet firm enormous power. "This is a more powerful monopoly than we've ever seen for access to 20th-century material," says Ms Moore of the University of Toronto. "We do not have a good track record in negotiating good prices with monopolies." Similar concerns led Harvard University to reduce its

participation in Google's project. Other librarians, however, regard the Google settlement as a good compromise, even if it is not perfect and does not address the criticisms that Mr Kahle and other internet types have with copyright law. "Brewster wants everything to be free," says Mr Courant of the University of Michigan. "So do I. But there are important trade-offs between what we collect and preserve and what we can make available."

Although the two projects take very different approaches—one idealistic, the other pragmatic—it may be that they will end up complementing each other. Librarians can and do work with both projects. And if things with Google go sour, libraries can always go elsewhere. "If Google's prices are too high, we can and will arrange with other players to re-scan the works. We still have the original source material," says Mr Courant. Consumers, likewise, are free to access public-domain books in either collection.

It may be that a lack of library funds, rather than Google, poses the biggest short-term threat to Mr Kahle's dream. Google covers the cost of scanning libraries' books. But to get into Mr Kahle's archive, libraries must either do their own scanning or pay the archive to do it. And, like everyone else, libraries are feeling the financial squeeze at the moment.

But Mr Kahle is taking a very long-term view. Universal online access to all knowledge may not be "a goal that is going to be finished in our lifetime," says Mr Kahle. "But if you pick a goal far enough out, people can align to it. I am not interested in building an empire. Our idea is to build the future." ■

Offer to readers

Reprints of this special report are available at US\$7.00 each, with a minimum of 5 copies, plus 10% postage in the United States, 15% postage in Mexico and Canada. Add tax in CA, DC, IL, NY, VA; GST in Canada.

For orders to NY, please add tax based on cost of reprints plus postage.

For classroom use or quantities over 50, please telephone for discount information.

Please send your order with payment by cheque or money order to:

Jill K. Aletha of Foster Printing Service
Telephone: 866 879 9144, extension 168
or e-mail: jillk@fosterprinting.com

(American Express, Visa and MasterCard accepted)